Meiotic crossover hotspots contained in haplotype block boundaries of the mouse genome

Liisa Kauppi*†, Maria Jasin‡, and Scott Keeney*

*Molecular Biology and ‡Developmental Biology Programs, Memorial Sloan–Kettering Cancer Center, 1275 York Avenue, New York, NY 10021

Edited by Kathryn V. Anderson, Sloan–Kettering Institute, New York, NY, and approved July 6, 2007 (received for review March 2, 2007)

Fertility requires successful chromosome segregation in meiosis, which in most sexual organisms depends on the formation of appropriately placed crossovers. The nonrandom genome-wide distributions of meiotic recombination events have been examined at the molecular level experimentally in yeast and by inference from linkage disequilibrium patterns in humans. Thus far, no method has existed for pinpointing sites of crossing-over on a genome-wide scale in an experimentally tractable animal whose genome size and complexity models that of humans. Here, we present a genomic approach to identify mouse crossover hotspots, based on targeting haplotype block boundaries. This represents a previously undescribed method potentially applicable to large-scale mouse hotspot identification. Using this method, we have successfully predicted the location of two previously uncharacterized crossover hotspots in male mice. As increasing amounts of single-nucleotide polymorphism data emerge, this approach will be useful for investigating the recombination landscape of the mouse genome.

meiosis | mouse strain | recombination | single-nucleotide polymorphism | allele-specified PCR

Much of the human genome consists of discrete chromosomal segments within which single-nucleotide polymorphisms (SNPs) are strongly associated (1–5). These segments, known as haplotype blocks, or linkage disequilibrium (LD) blocks, are 7- to 16-kb long on average, depending on the population (5). Haplotype block structure in humans is largely generated by the presence of preferred sites ("hotspots") for meiotic crossing-over, flanked by recombinationally inert DNA (6–9). The handful of human recombination hotspots analyzed at high resolution vary in their overall cross-over activity, but share the feature of having crossovers clustered within narrow, 1- to 2-kb regions (6, 7). High-resolution analysis of two hotspots in the mouse, initially identified from pedigree data, revealed similar properties (10, 11).

Extrapolating from extensive studies in yeast, it is likely that mammalian crossover hotspots are preferred sites for formation of the DNA double-strand breaks that initiate meiotic recombination (reviewed in refs. 12–14). However, although the existence of hotspots is well documented, the molecular mechanisms that control their activity are poorly understood. Availability of an experimentally tractable mammalian system for characterizing and manipulating hotspots is thus important.

Breakdown of LD, that is, disruption of haplotype block structure, reports on recombination that occurred during the history of the population. Thus, LD analysis is the current method for studying genome-wide fine-scale patterns of recombination in humans (reviewed in ref. 13). However, LD analysis is less suited for this purpose in mice, because of two features of the short and unusual population history of inbred laboratory strains. First, it has been argued that there is strong selection pressure to eliminate allelic combinations that cannot be tolerated in homozygous form (15). In this view, recombination events that break up favorable linkage groups reduce the fitness of progeny that inherit the recombinant haplotypes. Such recombination events would not leave their footprints in the haplotype structure because the recombinant progeny are selectively lost from the population. The net result is apparent suppression of recombination within certain genomic regions (refs. 15 and 16, but see also refs. 17, 18). Second, during the establishment of inbred strains, the number of meioses that create detectable crossovers is extremely limited, so only a relatively small number of crossovers have been captured in the haplotype structure. In contrast, the haplotype block structure in humans is the result of several hundred generations of meioses that occurred during the interbreeding of genetically diverse individuals.

Because of these unique features, haplotype blocks in inbred mice are typically orders of magnitude larger than in humans (18, 19). Thus, whereas haplotype blocks in the human genome are indicative of DNA that undergoes little or no recombination, haplotype blocks in mouse strains are uninformative about recombination activity within the blocks, and the number of boundaries between haplotype blocks is a gross underestimate of genome-wide crossover distribution. Nevertheless, it remains clear that observed disruptions in haplotype block structure originate from meiotic crossovers. It is therefore possible to identify sites where crossovers occurred during the derivation of the strains from ancestral *Mus musculus* subspecies (so-called historical crossovers, Fig. 1) by comparing SNPs in different inbred strains (18, 20).

In humans, it is estimated that >80% of crossovers occur at recombination hotspots (5, 6, 8, 21). Therefore, we hypothesized that most of the detectable historical crossovers in the mouse also occurred within hotspots and that many of these hotspots should be currently active in laboratory strains. We describe here a test of this hypothesis and a proof-of-principle demonstration of a way to identify novel recombination hotspots in the mouse.

Results

We used a two-tier strategy to identify and characterize crossover hotspots in the mouse, analogous to the approach of analyzing LD breakdown regions in humans (6). We first examined high-resolution haplotype structure to identify signatures of historical crossing-over. We then tested haplotype block boundaries for meiotic recombination hotspot activity by using allele-specific PCR to recover crossover molecules directly from sperm DNA.

Initially, we attempted to identify historical crossover sites based on published analyses of SNP distributions in pair-wise comparisons of several laboratory strains (19, 22). Specifically, we considered regions that showed a transition from SNP-rich to SNP-poor to be candidate hotspots (Fig. 1*a*). However, we were

Author contributions: L.K., M.J., and S.K. designed research; L.K. performed research; L.K., M.J., and S.K. analyzed data; and L.K. and S.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviations: ASO, allele-specific oligonucleotide; LD, linkage disequilibrium.

†To whom correspondence should be addressed. E-mail: kauppil@mskcc.org.

This article contains supporting information online at www.pnas.org/cgi/content/full/0701965104/DC1.

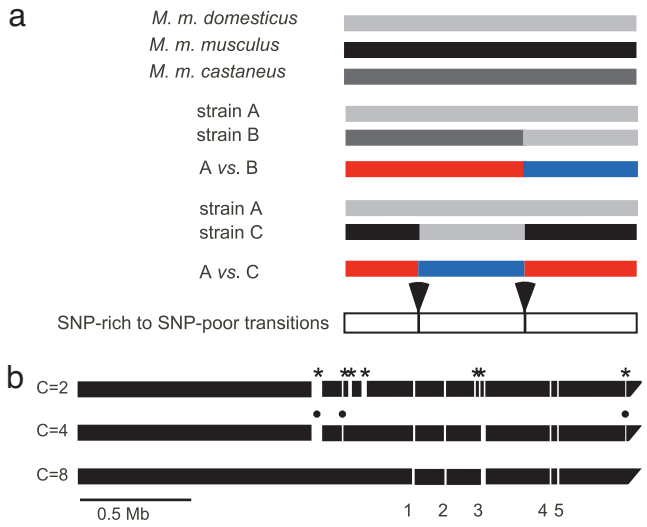

Fig. 1. Comparison of SNP distributions from several mouse strains reveals past crossover activity. (*a*) Schematic representation of SNP-rich to SNP-poor transition regions in a chromosomal domain. Three ancestral haplotypes are shown at the top in light gray, black, and dark gray. Any given genomic segment in a laboratory strain is a patchwork of these ancestral haplotypes (22), resulting in SNP-rich (red) and SNP-poor (blue) regions in a two-strain comparison. SNP-rich to SNP-poor transition regions should therefore represent sites of historical crossing-over, i.e., putative recombination hotspots. (*b*) Haplotype blocks in the first 2.5-Mb portion of the 4.8-Mb region on chromosome 1 examined by Yalcin *et al.* (20). Adapted and drawn to physical scale from ref. 20. Haplotype blocks at three transition costs (C) are shown. Boundaries between adjacent haplotype blocks at C = 8 are numbered. The region shown contains boundaries 1–5, of which boundaries 1, 2, and 3 were examined for hotspot activity in this report. A fourth boundary (boundary 9), not shown here, was also examined. Additional block boundaries are defined when the transition cost is lowered (C = 4, dots; C = 2, asterisks).

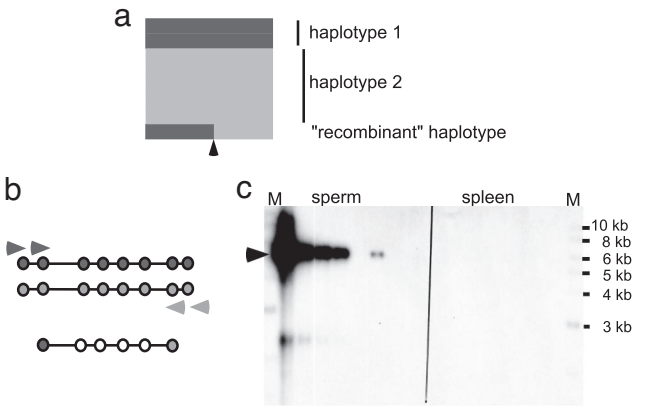

Fig. 2. Two-tier strategy to identify hotspots. (*a*) Historical crossover sites are located by examining haplotypes. Shown here is the haplotype structure around boundary 1 (from Fig. 1*b*) in eight strains, adapted from ref. 20. Each row represents one mouse strain. Strains A/J and C3H carry haplotype 1 (dark gray), strains BALB/c, C57BL/6J, DBA/2, RI, and RIII carry haplotype 2 (light gray), and AKR carries what appears to be a recombinant haplotype. The signature of historical crossing-over (boundary between two haplotype blocks) is indicated with a black arrowhead. (*b*) For putative recombination hotspot regions, two rounds of allele-specific PCR are carried out on batches of sperm DNA with multiple heterozygous SNPs (dark and light gray circles). Allele-specific primers are shown as dark and light gray arrowheads. Recombinant sperm DNA molecules are selectively amplified. The location of crossover breakpoints is mapped subsequently by typing internal SNPs. (*c*) An example of PCR products of pools of sperm and spleen DNA after the second round of amplification is shown from boundary 1 (fragment size ≈5.5 kb, arrowhead). The PCRs contained 20,000 input molecules each. Five of eight sperm DNA pools are positive for one or more recombinants; differences in signal intensity are likely due to stochastic differences in amplification efficiency. M, 1-kb ladder.

unable to identify crossover hotspots in this manner (data not shown). The main limitations of this approach are the relatively low density of resequencing amplicons for strain-to-strain comparisons (varying from one every ≈5 kb up to one every >200 kb) and, more importantly, the lack of unequivocal sharp SNP-rich to SNP-poor transition boundaries. A more recent study on a 4.8-Mb segment on chromosome 1 (20) alleviated both of these problems. In this work, densely spaced SNPs in eight strains were analyzed for haplotype block structure (Fig. 1*b*). A key feature of the computational algorithm used to define block boundaries was the inclusion of a parameter C (the cost of making a transition from one block to the next), with low C values favoring transitions (20). Thus, C controls the stringency of haplotype block boundary detection, with robust boundaries detectable even at high C values (Fig. 1*b*).

Direct Tests for Crossover Activity in Sperm DNA. We selected regions to assay for meiotic crossover activity based on the following criteria: the haplotype block boundary was computationally detectable at high transition cost (C = 8), the location of the boundary was defined to within ≈10-kb resolution (the limit of feasible long PCR), and sufficient numbers of SNPs were present on both sides of the putative hotspot in commonly used mouse strains to design nested allele-specific PCR primers (Fig. 2). As detailed below, we successfully amplified sperm-specific recombinant DNA molecules at two haplotype block boundaries (numbered 1 and 2 in Fig. 1*b*). PCRs seeded with spleen or liver DNA were used as a control for *in vitro* artifacts and for the meiosis-specificity of amplification products.

Crossovers at Boundary 1. Boundary 1 is located at 143.77 Mb on chromosome 1 [physical distance, National Center for Biotechnology Information (NCBI) assembly 33]. We performed crossover assays on DNA from a [C57BL/6J × C3H/HeJ] F_1 hybrid animal (from here on, referred to as [B6xC3H]), and recovered a total of 75 crossovers from 1,400,000 sperm DNA molecules from one mouse. Crossover breakpoints mapped within a 2-kb region, with a symmetrical distribution around the hotspot center similar to distributions in other hotspots (Fig. 3). Recombinant frequency was $\approx 5 \times 10^{-5}$ per sperm genome with a peak activity of 3 cM/Mb, which is ≈6 times the mouse genome average. No such crossover molecules were amplified by PCR of 1,300,000 somatic DNA molecules (frequency $< 0.8 \times 10^{-6}$; Fig. 2*c* and data not shown). Thus, this haplotype block boundary is indeed a hotspot for meiotic crossover formation. This hotspot, which we termed *M1*, is located in an ≈190-kb-long region devoid of known genes, although it is located in the putative second intron of a predicted transcript (GENSCAN00000381745).

Most DNA pools giving a sperm-specific PCR amplification (such as those shown in Fig. 2*c*) contained recombinant molecules that can be classified as "simple" crossovers, with a single, defined breakpoint between the parental haplotypes (e.g., Fig. 4*a*, molecule I). Less frequent were PCR products classified as "simple but mixed" crossovers: these PCR products had different parental haplotypes at either end of the amplicon, but had both parental alleles at one or more SNPs at the breakpoint (e.g., Fig. 4*a*, molecule II). These mixed PCR products were treated as two independent crossovers with distinct breakpoints (see Fig. 4*a*, II.i and II.ii for an example). It is possible that a subset of these PCR products arose instead from single crossover molecules containing unresolved heteroduplex DNA. At *M1*, however, the number of mixed PCR products (12) was identical to the number of pools of sperm DNA expected to contain two or more

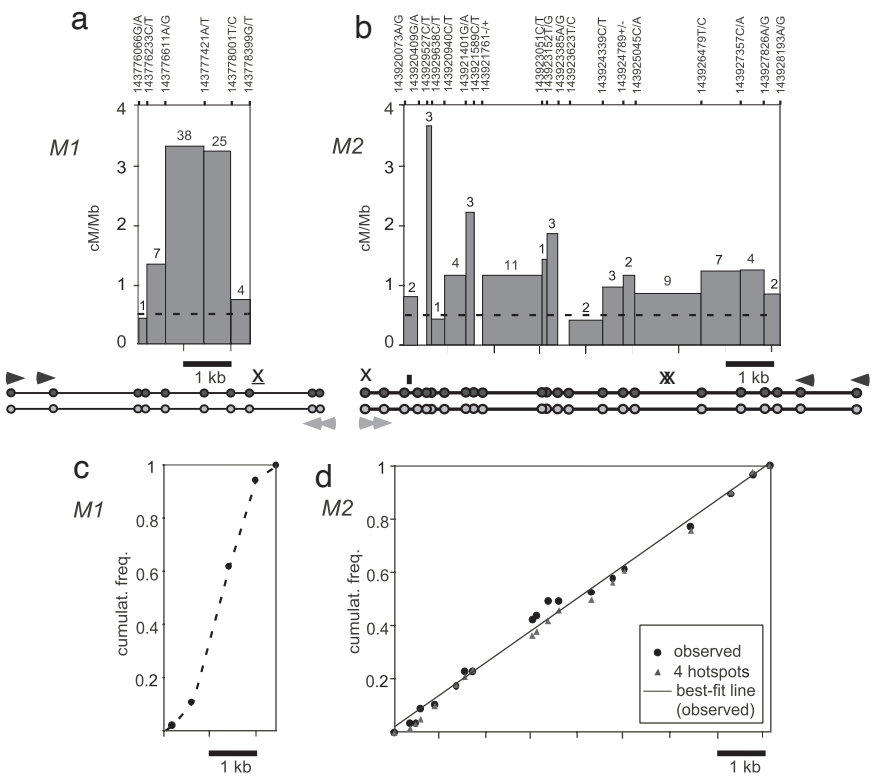

Fig. 3. Distribution of crossover breakpoints at hotspots *M1* and *M2*. (*a*) At the *M1* hotspot, allele-specific PCR was performed on sperm and somatic DNA from a [B6xC3H] F_1 animal. Positions of SNPs are shown as circles below the graph. Forward primers (dark gray arrowheads) are specific for SNPs on the C3H haplotype; reverse primers (light gray arrowheads) are specific for SNPs on the B6 haplotype. SNPs, which are named after their genomic location, are indicated as ticks above the plot. Bars indicate crossover activity in cM/Mb in each interval between heterozygous SNPs. Above each bar, the number of crossovers observed in that SNP interval is shown. We excluded end intervals (intervals between allele-specific selector SNPs used in secondary PCR, and first internal heterozygous SNP), because the lack of markers precludes distinction between parental bleed-through molecules and genuine crossovers. The dashed line indicates the average crossover rate on chromosome 1 in males (0.53 cM/Mb, from ref. 18). x, location of the heptamer CCCCCCT, closely related to the DNA sequence motif CCTCCCT reported to be overrepresented at human hotspots (21). (*b*) At the *M2* hotspot, allele-specific PCR was performed on sperm and somatic DNA from a [CZECHxC3H] F_1 animal. CZECH haplotype-specific forward primers are indicated as light gray arrowheads, and C3H haplotype-specific reverse primers are indicated as dark gray arrowheads. The location of the simple tandem repeat array is shown as a black rectangle below the plot. The structure of this array is $(AAAAC)_3(AAAGC)_{18}$ in C3H and $(AAAAC)_6(AAAGC)_2(AAAAC)_2$ in CZECH animals. x, locations of the CCTCCCT motif (21). (*c*) The cumulative distribution of crossovers at *M1* shows a sigmoidal shape, characteristic of other previously described hotspots. (*d*) For *M2*, the cumulative distribution of crossovers (filled circles) appears nearly linear. Also shown is the distribution for four hypothetical hotspots (triangles) distributed evenly across the interval (see *Materials and Methods* for details).

crossover molecules, based on the frequency of positive pools and assuming a Poisson distribution of recombinant molecules among the pools.

Crossovers at Boundary 2. Boundary 2 lies at 143.92 Mb on chromosome 1 (NCBI assembly 33). Crossover molecules were amplified from DNA of two [CZECHII/EiJ × C3H/HeJ] F1 hybrid animals (hereafter [CZECHxC3H]). (Note: there are not enough suitable polymorphisms across boundary 2 for allele-specific PCR between most common laboratory strains, so we set up a cross involving more distantly related strains. The CZECHxC3H cross was chosen because the F_1 animals are fertile, unlike most crosses involving CZECH and standard strains.) The assays yielded 57 crossovers from a total of 777,000 sperm DNA molecules, translating to a crossover frequency of $\approx 7 \times 10^{-5}$ per sperm genome. These recombinant molecules appeared to be sperm-specific, because none were amplified from 804,000 somatic DNA molecules (frequency $< 1.2 \times 10^{-6}$; data not shown). Unlike at hotspot *M1*, we observed an excess of PCR products containing simple mixed molecules (15 observed vs. 8 expected; data not shown). Possible reasons for this observation are addressed in *Discussion*. As with *M1*, these products were counted as containing two independent crossovers. Importantly, however, even if the excess PCR products were due instead to single crossovers with unrepaired heteroduplex DNA, counting them as two crossovers each had little effect (≤15%) on our estimate of overall crossover frequency (data not shown).

The distribution of crossover breakpoints in the 9-kb amplicon was markedly different from that at *M1* or from other previously described hotspots in mice and humans (Fig. 3*b*; compare cumulative distributions in Fig. 3 *c* and *d*). Instead of clustering in a narrow region of less than a few kilobases, crossovers were distributed rather evenly across the entire interval tested. These molecules are unlikely to be PCR artifacts, because they were not observed upon amplification of somatic DNA. Moreover, the breakpoint distribution of the simple but mixed recombinant molecules was similar to the distribution for simple crossovers, so inclusion of this species is not responsible for the unusual distribution of recombination events in this region (data not shown).

Although the crossover rate within individual SNP intervals in this region was only ≈2-fold above the mouse genome average, the events observed cannot be readily accounted for as randomly placed crossovers occurring outside hotspots. Such "background" crossovers are typically ≥20-fold less frequent than those observed at boundary 2 (refs. 6, 7, 10, 11, and our unpublished data). Therefore, boundary 2 contains a meiotic

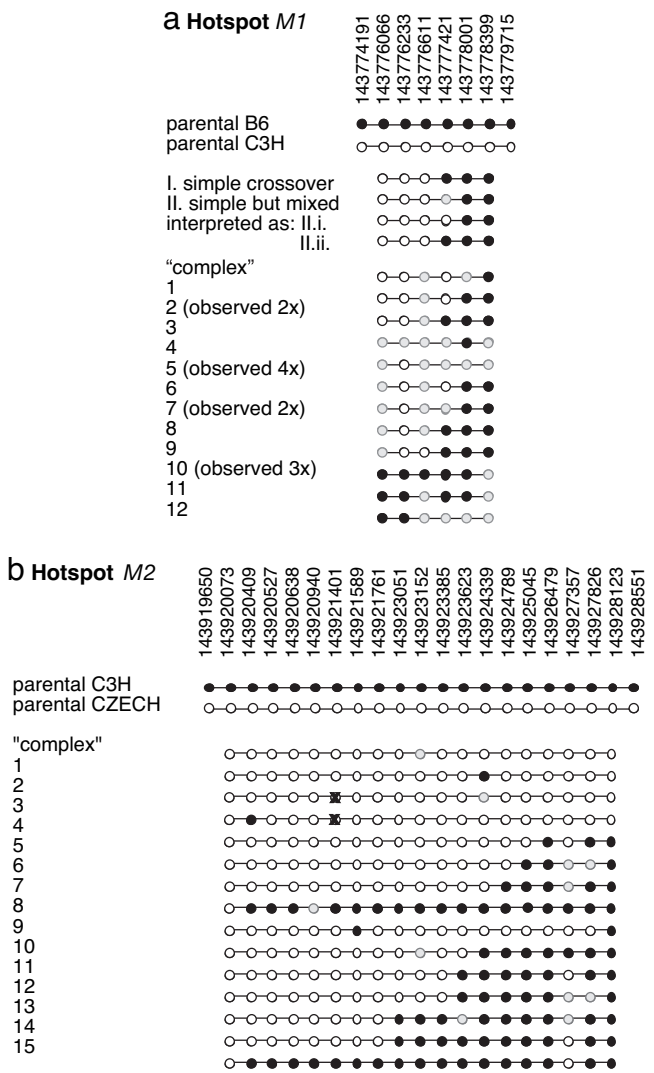

Fig. 4. Structure of ''complex'' amplified molecules in the *M1* (*a*) and *M2* (*b*) crossover assays. Parental alleles found at a given SNP locus are shown as black and white circles. Gray circles denote instances where ASO hybridizations gave a signal for both alleles. In *a*, examples of a simple crossover amplification product (I) and a simple crossover amplification product containing a mixed signal (II) are shown. These simple but mixed DNA pools (II) were assumed to contain two types of crossover, one of type II.i and one of type II.ii. The majority of crossovers at *M1* and *M2* fall into classes I and II. The remainder (19 DNA pools for *M1* and 15 DNA pools for *M2*) contained ''complex'' molecules; the structure of molecules in each pool is shown. Circles with X's across them in *b* are SNPs where no signal for either parental allele was detected upon ASO hybridization. The structure of our complex molecules is reminiscent of the ''mosaic'' and ''mixed strand'' containing recombinant molecules observed in *Mlh1*$^{-/-}$ male mice (25).

recombination hotspot that we termed *M2.* The *M2* hotspot lies ≈20 kb distal to the nearest gene, *Rgs2*.

''Complex'' Molecules at Boundaries 1 and 2. Most PCRs that gave a positive signal in the *M1* and *M2* assays had a structure that is explained most easily as one or two simple crossovers per pool of sperm DNA molecules. We did, however, also observe molecules with a more complex structure (Fig. 4). Overall, the occurrence of complex molecules was similar for both hotspots (19 complex vs. 75 simple molecules at *M1* and 15 complex vs. 57 simple molecules at *M2*). Each of the complex molecules seen at *M2* was unique (Fig. 4*b*), whereas some structures occurred more than once at *M1* (Fig. 4*a*). It is unlikely that the complex molecules originated solely from PCR artifacts, because we never observed similar molecules in control PCRs performed on somatic template DNA (data not shown). These could be genuine crossover molecules that contain unrepaired heteroduplex DNA or, in cases where only one base pair is involved, nucleotides misincorporated by *Taq* polymerase. Some could also derive from conversion tracts accompanied by crossover. Because of the ambiguity of their origin, these complex molecules were not included in crossover frequency calculations or in the maps of breakpoint distributions.

No Crossovers Detected at Boundaries 3 and 9. We also designed crossover assays across boundaries 3 and 9, located at 144.08 Mb and 145.28 Mb on chromosome 1 (physical distance, NCBI assembly 33), respectively. Large numbers of amplifiable molecules (>560,000 molecules of each progenitor haplotype, i.e., 6 μg of sperm DNA) were screened for both boundaries. We were unable to amplify any recombinant molecules from sperm DNA. Instead, PCR of both sperm and somatic DNA resulted in rare ($<1 \times 10^{-5}$ per DNA molecule) amplification of products containing nonrecombinant haplotypes, consistent with bleed-through of one progenitor haplotype due to mispriming by the allele-specific primers (data not shown). Boundaries 3 and 9 therefore do not represent currently active crossover hotspots, at least not in males of the inbred strains assayed (B6xC3H F1 hybrids).

Discussion

In this work, we identified two mouse crossover hotspots located at haplotype block boundaries. These studies serve as proof of principle that at least some sites of historical crossing-over are currently active recombination hotspots in laboratory mouse strains and that mouse genomic variation can be used to identify bona fide meiotic hotspots.

Unique Recombination Patterns at *M2*. Hotspot *M1* has characteristics that are in line with previously described hotspots in humans and mice, specifically its narrow (1–2 kb) and symmetric distribution of crossovers, consistent with a narrow zone for double-strand break (DSB) formation (6, 7, 10, 11). Hotspot *M2* is different, however, displaying a relatively even distribution of crossover breakpoints across a wide region. It is possible that crossovers at *M2* are initiated in a single, unusually broad (several kilobases) zone of DSB formation. Alternatively, the observed pattern could be caused by the presence of several ''typical'' hotspots closely spaced across the region (see Fig. 3*d*). Closely spaced clusters of DSBs have been observed in *Saccharomyces cerevisiae* [e.g., at the *CYS3* locus (23, 24)], although no information is available about the fine-scale distribution of crossover resolution sites in such regions. Clusters of crossover hotspots have also been described in humans, but these did not show an even distribution like that seen at *M2*. Instead, individual hotspots were separated from one another by DNA with substantially lower or no detectable crossover activity (6, 7).

It is also possible that the unusual distribution of crossover breakpoints at *M2* is a general feature of recombination occurring between relatively highly diverged DNA sequences. Allelic DNA sequence differences are 6.3-fold more common in the *M2* interval (99 differences in 8,980 bp, i.e., 1.1%, in [CZECHxC3H] F_1 hybrids) than in the *M1* interval (9 differences in 5,133 bp, i.e., 0.18%, in [B6xC3H] F_1 hybrids). The *M2* interval also contains simple tandem repeats which differ in number and structure between the two strains (see Fig. 3*b*).

At *M2*, unlike *M1*, we observed a greater number of PCR products containing simple but mixed molecules than expected from a Poisson distribution of single crossovers among the sperm DNA pools amplified. This result may indicate that some of the

sperm DNA pools contained single *M2* crossover molecules with persistent heteroduplex DNA. Similar molecules have previously been observed in mismatch repair deficient (*Mlh1*$^{-/-}$) mice (25). Because the animals examined in our study are mismatch-repair proficient, it is tempting to speculate that the heterozygosity in the simple tandem repeats at *M2* renders heteroduplex DNA more refractory to mismatch repair, allowing mismatches to persist from meiosis through spermiogenesis. There is ample precedent for poorly repaired mismatches escaping correction during meiosis in wild-type, i.e., mismatch repair proficient, fungi (for example, see refs. 26–29).

Finding Mouse Crossover Hotspots. The small number of previously known mouse hotspots was identified by pedigree analysis (10, 11, 30–32). Notably, however, pedigree analysis is limited in that it is able to identify only very active hotspots. Such extremely active hotspots may be unusual, with the norm in mammalian genomes more similar to "lukewarm" hotspots such as *M1* (and perhaps *M2*), displaying peak crossover activities several orders of magnitude lower than hotspots identified by pedigree analysis. On the entire human chromosome 12, for example, there are <10 recombination peaks predicted to be >50 cM/Mb, whereas there are hundreds of localized areas showing predicted recombination rates of 5–20 cM/Mb (ref. 21). The two-tier method described here has significant advantages over hotspot detection by pedigree analysis because of its greater spatial resolution and its ability to detect even relatively weak hotspots.

Interestingly, the entire 4.8-Mb region containing *M1* and *M2* is itself contained in a large chromosomal region with relatively low crossover activity, sometimes referred to as a recombination "desert" (18). Our results support the idea that recombination hotspots are ubiquitous in the entire genome, including regions that, on the large scale, can be classified as deserts.

Recombination hotspots cannot be extrapolated from the presence of genomic DNA sequence elements in any organism. The DNA sequence motif 5′-CCTCCCT, reported to be overrepresented at human hotspots (21), is found at or near *M2* three times, and a variant of it once near *M1* (Fig. 3 *a* and *b*). The significance of this motif remains elusive, however, because this sequence can also be found six times at boundary 3 and four times at boundary 9, where we detected no crossover activity. Hence, examining haplotypes from many mouse strains currently remains the best method for genome-wide *de novo* hotspot identification.

Our results suggest that haplotype block boundaries in inbred strains are a meaningful predictor of hotspot location. It seems unlikely that we would have obtained these results by chance, because hotspot *M1* was captured in its entirety and was located right in the middle of the 6.4-kb amplicon. Such precision in identifying hotspot position is expected for the method we applied in this work, and our success rate (0.5) is far higher than the probability we estimate of precisely locating at least two hotspots in four random attempts ($P = 0.006$). [The simple assumptions for this estimate are that 6% of total DNA is hotspot DNA (refs. 6 and 7 and the HapMap project), that this 6% is contained in 2-kb-wide hotspots and that, because of typical SNP density, a 6-kb amplicon is required to completely capture a 2-kb hotspot. We acknowledge that *M2* does not fit the model of 2-kb-wide hotspots, but there is no precedent of such wide crossover distributions, and therefore it is difficult to apply an alternative model.] In this context, it is worth noting a recent study by Tiemann-Boege *et al.* (8), in which a region of human DNA was scanned for crossover activity by using 17 nonoverlapping segments of ≈4 to 8 kb. Three hotspots were identified, but, importantly, none of them was precisely captured within a single amplicon. The genomic region has an overall recombination activity of 2.4 cM/Mb, which is ≈2-fold over human genome average and can thus be classified as a recombination hot domain, or "jungle." By comparison, the genomic region we analyzed has an overall recombination activity of 0.2 cM/Mb, ≈3-fold below genome average for mouse (17). It is reasonable to assume that hotspots are at least as numerous (if not more so) in hot domains vs. cold domains. Thus, the fact that none of the amplicons in the human study precisely captured a hotspot supports the conclusion that the probability of obtaining our results by chance is very low.

Our results show that a subset of haplotype breakpoints does contain active recombination hotspots, but limitations to this approach are also apparent. First, as discussed in Introduction, examining haplotypes in laboratory strains underestimates potential crossover positions because of limited numbers of meioses leading to the establishment of the strains and possible contribution of selective pressures. Extrapolating from human data [roughly one crossover hotspot per every 30–90 kb (6–8)], we would expect 50–160 recombination hotspots for the 4.8-Mb region assayed here, compared with 10 robust haplotype boundaries [for C = 8; at C = 2, 29 boundaries are detected (20)]. Inclusion of SNP data from larger numbers of inbred strains is one way to improve the power of this approach for defining haplotype block boundaries.

A second limitation is that not all positions of observed disruptions in haplotype block structure showed detectable sperm crossovers (namely, boundaries 3 and 9). Possible nonexclusive explanations for the absence of crossovers include hotspot activity in these regions existed in the past but has since been extinguished (33–36); hotspot activity is restricted to haplotypes different from the strain combinations we tested [haplotype-specificity has been observed e.g., at the Eβ hotspot (10)]; hotspot activity is female-specific (30); or these regions never contained a hotspot, and the recombinant haplotypes arose by rare, randomly placed exchanges.

Developing a collection of new crossover hotspots mapped at high resolution is crucial to begin to understand the rules of mammalian crossover distributions. A wealth of high-resolution SNP data from several strains has emerged recently (18, 20, 37, 38), and will continue to accumulate. These data will facilitate the identification of bona fide hotspots in many genomic regions, allowing the dissection of factors involved in mammalian crossover control.

Materials and Methods

DNA Amplification, SNP Discovery, and Genotyping. In addition to SNPs from the ENSEMBL mouse sequence database, SNPs were also discovered by resequencing DNA segments amplified by PCR from several mouse strains. Genomic DNA from mouse strains of interest was purchased from The Jackson Laboratories (Bar Harbor, ME) and amplified by PCR by using the buffer described previously (39), supplemented with 12 mM Tris base, 0.2 μM each primer, 0.03 unit/μl *Taq* polymerase and 0.012 units/μl Pfu polymerase. Primers are listed in supporting information (SI) Table 1. Newly discovered SNPs have been deposited into dbSNP (NCBI_ss#73418999–ss#73419013).

Crossover Assays. We designed and optimized allele-specific PCR primers as described (6). Sequences of allele-specific and universal primers, and PCR cycling conditions used in the crossover assays, are listed in SI Table 2. Caudal epididymides were dissected from adult mice, and DNA from mature sperm was extracted as described (40). Somatic DNA was extracted from spleen or liver. DNA was handled as described (41). Experiments conformed to relevant regulatory standards and were approved by the Memorial Sloan–Kettering Cancer Center Institutional Animal Care and Use Committee. To detect crossover molecules, multiple aliquots of sperm DNA, each containing input pools of 625–30,000 amplifiable DNA molecules (6.8–327 ng of

DNA) were amplified by long PCR (42) using allele-specific primers targeted toward outer selector SNP sites in recombinant phase ("primary PCR"). Somatic DNA from the same mouse was used at the highest DNA inputs as a negative control throughout. Primary PCR products were digested with S1 nuclease and were then reamplified ("secondary PCR") with nested allele-specific primers in recombinant phase. All secondary PCR products were subjected to a third round of PCR amplification with universal primers nested just inside the internal allele-specific primer sites. Crossover-positive reactions were counted for each input pool size, and crossover rates were calculated by using Poisson corrections for multiple crossover molecules in the positive reactions. All crossover-positive reactions were transferred onto dot blots, and crossover breakpoints were mapped by allele-specific oligonucleotide (ASO) hybridization. Further details of methods for detecting and mapping crossover molecules have been published (6, 43). ASO sequences are listed in SI Table 3.

Modeling Crossover Hotspots at Boundary 2. To assess whether the distribution of crossovers at hotspot *M2* could result from several weak hotspots scattered across the interval, we plotted distributions expected for four hypothetical hotspots. These hotspots were spaced evenly across the interval, each was 1.5-kb wide with symmetrically distributed crossovers, and each displayed identical crossover activity. The total number of observed crossovers (57) across the interval was divided evenly between the four hypothetical hotspots.

We thank Alec Jeffreys, Francesca Cole, and members of the S.K. laboratory for comments and suggestions on the manuscript. This work was supported in part by National Institutes of Health Grant HD 40916 (to M.J. and S.K.) and the Helsingin Sanomat Centennial Foundation (L.K.). S.K. is a Scholar of the Leukemia and Lymphoma Society.

1. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins, J, DeFelice M, Lochner A, Faggart M, *et al.* (2002) *Science* 296:2225–2229.
2. Phillips MS, Lawrence R, Sachidanandam R, Morris AP, Balding DJ, Donaldson MA, Studebaker JF, Ankener WM, Alfisi SV, Kuo FS, *et al.* (2003) *Nat Genet* 33:382–387.
3. Ardlie KG, Kruglyak L, Seielstad M (2002) *Nat Rev Genet* 3:299–309.
4. Dawson E, Abecasis GR, Bumpstead S, Chen Y, Hunt S, Beare DM, Pabial J, Dibling T, Tinsley E, Kirby S, *et al.* (2002) *Nature* 418:544–548.
5. The International HapMap Consortium (2005) *Nature* 437:1299–1320.
6. Jeffreys AJ, Kauppi L, Neumann R (2001) *Nat Genet* 29:217–222.
7. Jeffreys AJ, Neumann R, Panayi M, Myers S, Donnelly P (2005) *Nat Genet* 37:601–606.
8. Tiemann-Boege I, Calabrese P, Cochran DM, Sokol R, Arnheim N (2006) *PLoS Genet* 2:e70.
9. Greenawalt DM, Cui X, Wu Y, Lin Y, Wang HY, Luo M, Tereshchenko IV, Hu G, Li JY, Chu Y, *et al.* (2006) *Genome Res* 16:208–214.
10. Yauk CL, Bois PR, Jeffreys AJ (2003) *EMBO J* 22:1389–1397.
11. Guillon H, de Massy B (2002) *Nat Genet* 32:296–299.
12. de Massy B (2003) *Trends Genet* 19:514–522.
13. Kauppi L, Jeffreys AJ, Keeney S (2004) *Nat Rev Genet* 5:413–424.
14. Petes TD (2001) *Nat Rev Genet* 2:360–369.
15. Petkov PM, Graber JH, Churchill GA, DiPetrillo K, King BL, Paigen K (2005) *PLoS Genet* 1:e33.
16. Petkov PM, Graber JH, Churchill GA, DiPetrillo K, King BL, Paigen K (2007) *PLoS Biol* 5:e127.
17. Shifman S, Bell JT, Copley RR, Taylor MS, Mott R, Flint J, Williams RW (2007) *PLoS Biol* 5:e128.
18. Shifman S, Bell JT, Copley RR, Taylor MS, Williams RW, Mott R, Flint J (2006) *PLoS Biol* 4:e395.
19. Wiltshire T, Pletcher MT, Batalov S, Barnes SW, Tarantino LM, Cooke MP, Wu H, Smylie K, Santrosyan A, Copeland NG, *et al.* (2003) *Proc Natl Acad Sci USA* 100:3380–3385.
20. Yalcin B, Fullerton J, Miller S, Keays DA, Brady S, Bhomra A, Jefferson A, Volpi E, Copley RR, Flint J, *et al.* (2004) *Proc Natl Acad Sci USA* 101:9734–9739.
21. Myers S, Bottolo L, Freeman C, McVean G, Donnelly P (2005) *Science* 310:321–324.
22. Wade CM, Kulbokas, EJ, III, Kirby AW, Zody MC, Mullikin JC, Lander ES, Lindblad-Toh K, Daly MJ (2002) *Nature* 420:574–578.
23. de Massy B, Rocco V, Nicolas A (1995) *EMBO J* 14:4589–4598.
24. Vedel M, Nicolas A (1999) *Genetics* 151:1245–1259.
25. Guillon H, Baudat F, Grey C, Liskay RM, de Massy B (2005) *Mol Cell* 20:563–573.
26. Kramer B, Kramer W, Williamson MS, Fogel S (1989) *Mol Cell Biol* 9:4432–4440.
27. Nag DK, White MA, Petes TD (1989) *Nature* 340:318–320.
28. Hillers KJ, Stahl FW (1999) *Genetics* 153:555–572.
29. Kirkpatrick DT, Petes TD (1997) *Nature* 387:929–931.
30. Shiroishi T, Koide T, Yoshino M, Sagai T, Moriwaki K (1995) *Adv Biophys* 31:119–132.
31. Buchner DA, Trudeau M, George AL Jr, Sprunger LK, Meisler MH (2003) *Genomics* 82:452–459.
32. Kelmenson PM, Petkov P, Wang X, Higgins DC, Paigen BJ, Paigen K (2005) *Genetics* 169:833–841.
33. Ptak SE, Hinds DA, Koehler K, Nickel B, Patil N, Ballinger DG, Przeworski M, Frazer KA, Pääbo S (2005) *Nat Genet* 37:429–434.
34. Winckler W, Myers SR, Richter DJ, Onofrio RC, McDonald GJ, Bontrop RE, McVean GA, Gabriel SB, Reich D, Donnelly P, *et al.* (2005) *Science* 308:107–111.
35. Wall JD, Frisse LA, Hudson RR, Di Rienzo A (2003) *Am J Hum Genet* 73:1330–1340.
36. Neumann R, Jeffreys AJ (2006) *Hum Mol Genet* 15:1401–1411.
37. Zhang J, Hunter KW, Gandolph M, Rowe WL, Finney RP, Kelley JM, Edmonson M, Buetow KH (2005) *Genome Res* 15:241–249.
38. Frazer KA, Wade CM, Hinds DA, Patil N, Cox DR, Daly MJ (2004) *Genome Res* 14:1493–1500.
39. Jeffreys AJ, Neumann R, Wilson V (1990) *Cell* 60:473–485.
40. Zhang Y, Monckton DG, Siciliano MJ, Connor TH, Meistrich ML (2002) *Mutat Res* 516:121–138.
41. Jeffreys AJ, Tamaki K, MacLeod A, Monckton DG, Neil DL, Armour JA (1994) *Nat Genet* 6:136–145.
42. Cheng S, Chang SY, Gravitt P, Respess R (1994) *Nature* 369:684–685.
43. Jeffreys AJ, Ritchie A, Neumann R (2000) *Hum Mol Genet* 9:725–733.